AHRC ICT Methods Network Workshop

SPACE AND TIME: METHODS IN GEOSPATIAL COMPUTING FOR MAPPING THE PAST

WELSH E-SCIENCE CENTRE, CARDIFF UNIVERSITY, 10 – 11 OCTOBER 2007

SECTION 1: OVERVIEW

Stuart Dunn, King's College London
Leif Isaksen, University of Southampton

This report summarizes an AHRC ICT Methods Network[1] workshop held on 23 and 24 July 2007 at, and jointly supported by, the e-Science Institute[2] in Edinburgh.

Entitled *Space and Time: Methods in Geospatial Computing for Mapping the Past*, the event's aim was to:

- Take stock of current developments in advanced geospatial technologies, particularly but not exclusively in the domains of history and archaeology;
- Assess the research questions facing practitioners in those fields;
- Consider future directions.

Format of the workshop and report

The workshop, which was held in conjunction with a three-part lecture under the auspices of the Arts and Humanities e-Science Theme, *Aspects of Space and Time for Humanities e-Science*[3] fell into three sessions on key areas: **Scale**, **Heterogeneity** and **Standards and Metadata**. Each session was introduced by a keynote lecture, respectively Vince Gaffney, Tom Elliott and William Kilbride, who based their lectures on previously circulated position papers. The sessions themselves comprised short presentations from researchers and expert practitioners, followed by general discussion. An expert rapporteur (respectively William Kilbride[4], David Wheatley and Stephen Stead) summed up each session, and provided a report.

Key Outcomes

Throughout the two days, there was significant disagreement in a number of academic and strategic areas, and many of these remained at the end. In this regard, it is as interesting to reflect on what was *not* discussed as what was. For example, as William Kilbride notes in his report on the Scale session, there 'was little discussion

1 http://www.methodsnetwork.ac.uk (last accessed 14/2/2008)
2 http://www.nesc.ac.uk/esi (last accessed 13/2/2008)
3 http://www.nesc.ac.uk/esi/themes/theme_06/ (last accessed 30/11/2007).
4 Professor Gary Lock of Oxford University initially agreed to act as rapporteur for the Scale session, but was prevented from attending due to unavoidable last-minute circumstances. We express our grateful thanks to Dr. Kilbride for stepping in at the last minute.

of the continuing rise of high performance computing', or of the intimately related question of the integration of e-science technologies in geospatial computing. Furthermore, the issue of the *integration* of space and time was not explored in anything like the depth the subject merits: indeed the discussion tended to revert to the more familiar pastures of purely spatial matters.

It was, in some ways, a disappointment that no clear plan of action or agreed vision emerged, but this is certainly not to criticise anyone involved, or to understate the value of the debate that was held. On the contrary, the diversity of opinions expressed, and the vibrancy of the debate in all the sessions are, we believe, indicative of a still-emerging field that has an established set of (quantitative) tools, yet is still defining its intellectual parameters, and is feeling its way into new and unfamiliar social, technological and intellectual areas. This report attempts to provide an overview of the current state of these parameters under the key thematic areas that emerged.

Despite the lack of a definitive and agreed next step, we consider the event a success in that, based on the papers and discussion, we have been able to identify four key areas of priority for geospatial computing for mapping the past:

- Large scale modelling and visualization
- Web 2.0 and spatial data (infra)structures
- Documentation
- Integrating time with space

These seemed to us to be the main issues that emerged, and as organizers, we would like to see further meetings focusing on these areas. These would also be good headings for small exemplar projects to develop ideas, and their applicability to the humanities, in much more detail.

The first part of this document has been prepared by the organizers as an overview of the whole discussion, and as an attempt to draw together the various strands that emerged across the whole two days under the headings above. The position statements and reports provided by the keynotes and rapporteurs are provided in the second section and, taken together, provide focused expert perspectives on the three key areas. We hope the discussions will continue on the Arts-Humanities.Net community portal, available at http://www.arts-humanities.net/event/43. A key point to emerge in the Standards and Metadata discussion is that an emphasis is needed on *processes*, and not just *objects* – this document is part of a process, and makes no claim to define anything other than the current moment in that process. It is essential also to note that each section of this report, including this one, represents **only the views of the individual authors,** and not in any way a communiqué of the workshop, or a consensus of all the participants.

KEY STRATEGIC AREAS

1. Large Scale Modelling and Visualization

Agent-based Modelling (ABM) is a type of computational modelling in which programmatic 'agents' are given a rule set and framework in which to operate and then made to interact in order to observe how they behave. In this way researchers attempt to model real world behaviour to see how large scale patterns 'emerge' from smaller ones. Doubts have occasionally been raised about this approach due to a (perhaps) unfounded perception that ABM attempts to make 'strong' claims about its

results. Not to the degree of individual agent histories, but to the extent that conclusions might be drawn that, "X and Y are the factors which led to outcome Z and this is how they interoperate". Criticisms range from the problem of equifinality (i.e. a variety of different processes might potentially lead to Z, and it is not possible to ascertain which), to whether such massive generalizations over complex systems can provide any meaningful results at all.

ABM emerged as a particular topic of interest in the Scale session, and provided a concrete example of how the large scale predictive methodologies can be deployed. It was viewed as a potential area to test current and future computational capabilities and capacity; and reflection on the assumptions underlying modelling approaches. However, the more general possibilities afforded by e-science and large scale e-infrastructure for the introduction of High Performance and High Throughput Computing in the manipulation and visualization of geospatial data must also be addressed in the (near) future.

The presentations by Tony Wilkinson and Mark Lake in the session on scale provided specific perspectives. Wilkinson presented some of the methodology and results of the Modelling Ancient Settlement Systems (MASS) Project[5]. The work, undertaken by a group of institutions between the universities of Durham and Chicago, and Argonne National Laboratory, has created an extremely sophisticated model of early Mesopotamian tell sites based on what we know of their populations, locations and internal dynamics. Some of the most interesting conclusions are that the effects of crises such as drought on such societies may play out in a variety of complicated ways that substantially challenge simpler 'decline and fall' type hypotheses. Such results can then be used to look for tell-tale signatures that can be looked for in the archaeological record.

Mark Lake then gave a more general critique of the ABM process, based in part upon his own previous work in Islay[6]. He made a strong case that, in the hands of a reflective and reflexive research community, it can be a valuable tool. When used as an experimental 'laboratory', it may throw up all manner of interesting insights and possibilities which might not otherwise occur to the researcher. The main dangers arise when the process is undertaken without a critical appraisal of the quality and of source data in comparison to the nature of the questions being asked of it (see the sections in the report by Kilbride and Stead for further discussion of this point).

These observations have parallels in other communities based around other advanced ICT methods (such as those working with Network Analysis for example). Where there is a fairly small user community, all using related but distinct approaches, there is frequently a danger of 'over-selling' the method in attempt to establish its utility. In all of these cases it is important to critique our own and each others' methodologies in order to establish firm theoretical ground. If we don't, they may unnecessarily become passing fads.

---

5 http://oi.uchicago.edu/OI/PROJ/MASS/introduction.htm (last accessed 14/2/2008)

6 Lake, M. W. 2000: MAGICAL computer simulation of Mesolithic foraging on Islay. In: S.J. Mithen, Editor, *Hunter–Gatherer Landscape Archaeology: The Southern Hebrides Mesolithic Project, 1988–98*, *Archaeological Fieldwork on Colonsay, Computer Modelling, Experimental Archaeology, and Final Interpretations* **vol. 2**, The McDonald Institute for Archaeological Research, Cambridge (2000), pp. 465–495.

2. Web 2.0 and Spatial Data (Infra)structures

The main focus of Tom Elliott's paper, and the following discussions, was how broader communities can be involved in geospatial research. This is a topic of considerable broader interest as the Web 2.0 agenda develops, and the concomitant 'Geospatial Web' emerges.[7] In particular the workshop considered the tension (described from time to time as a 'religious war') between 'bottom up' approaches and 'top down' ways of managing and delivering geospatial data.

A directly related issue is that of folksonomies as contrasted with ontologies**,** and how the process of structuring geospatial data is most appropriately carried out. This in turn links to a more established area of geospatial research: semantics. Should semantic systems be designed centrally by a small group of experts and distributed, or should these be community driven? What are the roles of organizations such as the Archaeology Data Service and EDINA?

Much of the discussion picked up the well-rehearsed observation that, not only has basic geocomputing become available to the public, but that in a very short space of time it has become hugely influential in the Web (and thereby cultural) sphere. Huge quantities of spatial data are becoming available all the time, often created by people with their own implicit or explicit agenda. Whilst professional archaeologists shouldn't turn their backs on traditional methods of dissemination, we must also utilize common dialects (such as KML and GeoRSS) if we are to participate in that wider public dialogue. Nobody owns the past, and our role as academics and professionals can only influence, not direct, other people's views, but these communities have a responsibility to ensure that it is important to make sure that we don't just end talking to ourselves.

The understanding and representation of any information in archaeology is dependent on common semantic frameworks and vocabularies, and is directly related to the preceding discussion of ownership of the past: who creates and owns (these are not necessarily the same) such frameworks and vocabularies? This has been the subject of threads on numerous email discussion fora in the past few months; a debate that was reflected by vigorous discussion in the workshop. There was recognition that standards must be robust, sustainable and trusted, with one participant noting on the blog:

> 'I absolutely agree with [the] assertion of the importance of common standards. There is no excuse for inventing yet another schema and adding to the informati-chaos that is out there (although the Google-ettes would argue that with the now almost infinite power of data-mining the folksonomized chaos is a good thing). What happens, however, when your data model is more complex than the most appropriate standard in the field? Do we dumb down our data? Do we create a dumbed-down serialisation of our data in the standard, while also using this as a pathway into our more complex and useful data? Do we try and modify the existing standards or invent new ones that are better and more appropriate?'[8]

---

[7] See for example Scharl, A. and Tochtermann, K. (eds.) 2007: The Geospatial Web: how geobrowsers, social software and the Web 2.0 are shaping the network society. Springer-Verlag, London.

[8] http://www.arts-humanities.net/mapping_past/212?page=1 (last accessed 143/2/2008)

A key aspect to emerge from the workshop is that this well-known dichotomy takes on a new meaning when research is conducted in the kind of mash-up based virtual research community described by Tom Elliott in his keynote talk. Elliott noted that UNC's Pleiades project[9], the (digital) successor to the widely respected, and heavily relied upon (paper) Barrington Atlas, could not have assembled the large, diverse, and globally dispersed team of professionals that it has using the traditional model used by the Barrington itself: the contributions of such a community to even a single project simply cannot be regulated and codified according to a pre-formulated ontological arrangement. Indeed this would negate the point of having such a freeform project structure. This, therefore, is an approach which very naturally lends itself to Web 2.0-style community driven tagging. But this too brings challenges. In the discussion following Elliott's paper, it was noted that this calls into question the very notion of 'encoding place'. There are potentially many different things that can comprise a 'place'. In one example Elliott gave, a picture of a camel had been uploaded to Flickr and associated by user tagging with the site of Aspendos, Turkey. Although this photograph may not be of immediate relevance to most researchers concerned with the history of Aspendos in classical antiquity, this picture is nonetheless a valid component Aspendos as an 'encoded place'. It is, in other words, part of the site's 'narrative'. In the sense that this provides a very rich and deep body of heterogeneous data with common associations provided by different users *wherever they are coming from*, this could be an example of how 'folksonomized chaos is a good thing'.

In parallel with these incoming and outgoing data streams run sets of standards, which regulate and map them. As Leif Isaksen pointed out in his paper, spatial information can be both textual and visual. Dealing with the textual component, whether in narrative historical text, archaeological reports, tables associated with a map, of even labels describing entities on a representation, must be founded upon the use of semantic standards, and it is critical to be aware of where those standards come from. In this context, the workshop recognized that there is a distinction between 'top down' approaches to standards which are imposed from above, and 'bottom up' approaches, where communities form their own sets of standards: a tension, in other words, between formal taxonomic and ontological systems which are set by some constituted authority or authorities, and unconstrained, or at least less-constrained, user-created 'folksonomy' tagging systems.

At the workshop, this well-rehearsed debate took two unexpected and interesting directions. In his summing up of the heterogeneity section, David Wheatley asked if the consequences of 'getting things wrong' were as dire in archaeology as much debate in this area assumes; and if there are grounds for the community to be seriously worried about what users will do with generally-available data. The workshop recognized that tags represent what is important to the user, rather than adherence to some agreed standard, and that this might assume a different order of significance in, for example, medical science. However, Vince Gaffney pointed out that much of his own field experience in the former Yugoslavia, where many instances of armed conflict had focused on sites of cultural and historical importance, provides a powerful demonstration that archaeological interpretation can have serious consequences in politically volatile areas.

9 http://pleiades.stoa.org/ (last accessed 143/2/2008)

3. Documentation

This valuable discussion centred on the need to track the lifecycle of geospatial content, from concept to output to reuse at a technical, content and metadata level. Clarity of where data has come from and how it has been created is as important as the data itself, a point forcefully made by Stephen Stead in his summing up of the Standards and Metadata session. This may appear an obvious point, but it is in fact a far broader scope of what is often meant by 'documentation' in digital preservation, where the term is often assumed to refer only to (static) data collections and objects. Broad agreement was reached that standards must be used to structure geospatial content, not constrain it.

Stead noted in his presentation (and reiterates below) that technology evolves rapidly, and standards frequently evolve at a slower pace. This leads to significant problems, and reflects the need for standards to describe research, and to reflect best practice, rather than to constrain it. This key principle guided much of the discussion on documentation: a point frequently made by many of the speakers, and upon which most participants were unanimous, is that standards are not neutral. Documentation defines the thing being documented.

A further key point was made early on, in Femke Reitsma's presentation in the preceding *Aspects of Space and Time* lecture, that it is an easy matter to document a particular object in space at a particular time, and then to document it again at a future moment, to quantify the length of time elapsed and to compare the object in its first state with its second. However, it is far more problematic to describe the *process* by which one led to the other. This is a critical issue for dealing with space *and* time in the humanities generally (and not just in historical and archaeological applications), and links with other important theoretical areas: for example, it could be argued that taking a predictive approach to individual decision making – as in agent-based modelling – might allow the archaeologist to reconstruct how a society developed from one condition into another. But the caveats critically assessed by William Kilbride in his rapporteur paper would undoubtedly remain.

This analysis of the documentation of process resurfaced in the Standards and Metadata discussion as a debate on the importance of objects **and** the significance of the relationships between objects. It was noted that all artefacts of cultural heritage have attributes of some sort. For example, a vessel from an excavation will be describable in terms of decoration, shape, material, date (an attribute often derived from others by typological placing) and so on; but knowledge about the cultural context of that vessel can only be (re)constructed on the basis of knowledge about other vessels.

Finally, as highlighted by the EDINA presentations of James Reid and Guy McGarva, effective documentation is essential to relevant IPRs, and for enabling proper citation and credit. This relates back to the folksonomy/ontology question: *what* is being cited, and by *whom* was it created are not simple questions in 'neogeography'. The issue of database rights versus copyright was also discussed in this regard: as Stead notes in his report, few concrete conclusions were drawn, but this is an area which requires further investigation.

4. Integrating time with space

It is a testament to the importance of this area that it is clearly relevant to all three areas of discussion. As Gaffney notes in his keynote paper below: 'there must be some concern not with what scaled data represents, but what it signifies, adds or subtracts from our interpretative schema.' In the first session it was noted that, while scale is most normally associated with spatial information, temporal scales are just as important to any representation of data; the second session noted that datasets change through time as they are added to and edited, and the third noted that metadata and standards are needed to describe time-based data as well as spatial information. However, one of the most interesting critiques of the workshop, and a common feature of the participant feedback, was that it had dwelled rather too heavily on the well-trodden topic of spatial representation, and not enough on time. As organizers, we are inclined to agree that there is much more that could be done in this field, and the comparative lack of papers focusing on the issue was representative of this fact.

Although Ian Johnson's TimeMap software is still more or less the only application to take the integration of time and space head-on (and we were lucky to have his presence in cyberspace, especially given the problems of time caused by location in Australia), alternative Semantic Web based approaches, such as those developed for the SIMILE project at MIT, are also emerging. In both cases, the real challenges seem less to be in what manner to present information using either temporal metaphors in the form of moving images, or spatial metaphors in the form of timelines – this has been substantially addressed by, for example Google Earth's *TimeSpan* and *TimeStamp* functions. The central challenge is how can we create a data-matrix which is rich enough to provide remotely even coverage?

This issue was addressed to a limited degree by Leif Isaksen's paper on Truth and Falsity in Heterogeneous Data, which argued that many (indeed most) geospatial representations conceal the limitations of their underlying dataset. This means that uneven temporal coverage is not so apparent to the end user, but this can lead to false assumptions about the compiler's intent (which would be more obvious if displayed temporally as well as spatially). More work needs to be done in classifying the ways in which data has been compiled if we are to create 'intelligent' GIS systems which can suggest suitable manners by which to frame the accessible information, such as it is (this point is further discussed in general terms in the Standards and Metadata contributions below from Kilbride and Stead).

SECTION 2: SESSION REPORTS

SESSION 1: SCALE

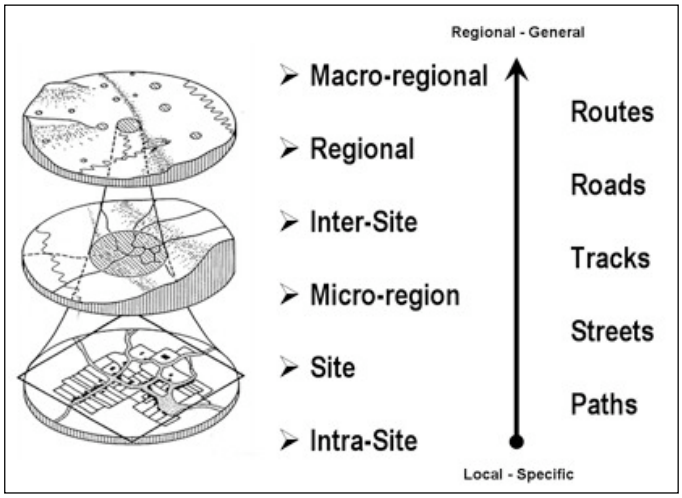
SCALE: A MEAGRE PRODUCT OF REALITY

Vince Gaffney, University of Birmingham

The title of this paper originates in a quote by Friedrich von Schiller who exhorted us to 'measure not by the scale of perfection the meagre product of reality'. It is actually a quote that I have used before in a published paper on scale but, perhaps, the sentiment of the quote deserves further consideration within the context of a wider

discussion on the significance of scale within the humanities[10]. In attempting to do so now I would acknowledge that Schiller would have been appropriately reassured by the certainty that archaeology and the humanities more generally are only rarely troubled by the problems of perfection. Moreover, within the context of archaeological analysis, at least, it is also a moot point whether many of our digital products are, or should be considered, a product of reality. Such observations may, perhaps, be dismissed as trite but the underlying point is a serious one and impinges upon any discussion associated with the issues of scale. Within the historical studies, the substance of our study, past action and meaning, can never be engaged with directly. This is in contrast to any metric description of an existing material residue, its associated physical context or the larger natural environment. Consequently, any aspect of representation carries significant implications in relation to how we understand or interpret our data. In spatial terms, an understanding of scale and, by association, the allied concept of resolution should therefore be central to our research.

Archaeology, in common with most humanistic studies, is inherently a spatial discipline. In empirical terms its practitioners are usually comfortable with discussion relating to the two factors discussed here, scale and resolution, which are encountered almost continuously during the process of data collection, manipulation and display. The definitions provided in the ADS GIS Guide to Good Practice are as useful as any in establishing a primary point of reference[11]. Here we learn that scale is the ratio of the distance measured on a map to that measured on the ground between the same two points whilst resolution is the smallest distance that can be usefully distinguished on a map with a given scale. Whilst acknowledging the significant related issues of precision and accuracy, a prime quality of both scale and resolution is that, together, they represent significant abstractions of a potentially experienced reality. In cartographic terms the process of abstraction is often associated with map generalization: a procedure associated with a considerable technical literature but which, in empirical terms, is almost invariably concerned with problems of direct representation. Whilst this is undoubtedly of value when ordering our data, this also presents us with significant problems. If we regard scale as being



10 Gaffney, V. and Gaffney, C. 2006: No further territorial demands: on the importance of scale and visualization within archaeological remote sensing. In "From Artefacts to Anomalies: Papers inspired by the contribution of Arnold Aspinall. University of Bradford 1-2 December 2006. (http://www.brad.ac.uk/archsci/conferences/aspinall/ and http://www.brad.ac.uk/archsci/conferences/aspinall/presentations/Gaffney&Gaffney.pdf)

11 http:/ **Figure 1. The significance of scale (modified from Roberts 1996, figure 2.2, and reproduced in Gaffney and Gaffney (forthcoming).**

a derived, simplified and therefore manufactured value, there must be some concern not with what scaled data represents but what it signifies, adds or subtracts from our interpretative schema. In other words, what is excluded from a data set as a consequence of scale may not be as important as the significance of the derived interpretation.

We can explore some of these matters through a consideration of the issues associated with the relatively innocuous interpretation of routes and communication – a mainstay of landscape analysis in many disciplines. As we move between larger and smaller scales the nature and interpretative significance of routes varies accordingly. Here, tracks, streets or roads may have an explicit or formal existence but routes, for instance, may be almost conceptual in their scale of representation or significance. The nature of the Silk Route, for instance, is a highly contentious issue. Another characteristic of these phenomena is that the behavioural significance of routes may also vary, not only according to the subject of study, but also according to the scale at which any analysis may be undertaken. Different processes may operate at different scales and, therefore, scale or resolution is not an objective optic on the past or indeed the present[12].

These issues can become more problematic if one considers the associated qualities of routes. Accessibility, for instance, is a more general characteristic measuring the relative ease or difficulty with which one may cross or communicate across or within an area of land. Communication may also be a product of a route or the general accessibility of a landscape and it may, equally, take static or mobile forms. Issues affecting these qualities, and one might readily acknowledge discussion related to the use of viewsheds in archaeology, are equally affected by scale and resolution. Behavioural issues associated with the choice of scale are therefore substantial and these may become increasingly complex when temporal factors are incorporated into our analyses.

Having acknowledged the significance of scale it is important to consider, at least, whether the choices we make in relation to scale, and the undoubted limitations that arise as a consequence of such decisions, are actually required. There must be an argument that in some instances the increasing resolution of our spatial description must approach, if not perfection, at least adequacy on occasion. Laser metrical survey, whether as air-based LiDAR or ground-based 3D surface scanning, might be considered in this light, as may some aspects of remote sensing[13]. The increasing capacity of parallel or Grid networks to represent or analyse such data might also prompt such an observation[14]. Unfortunately, whilst these achievements are significant in their own terms, the adequacy of measurement does not presume interpretative significance. We cannot necessarily finesse a path to understanding through an increasing resolution of measurement or enhanced access to computational power[15].

---

12 Gaffney, V. and Gaffney, H. Forthcoming. Modelling routes and communications. In Külzer A. (Ed.) Handelsgüter und Verkehrswege. Commodities and Traffic Routes. Aspekte derWarenversorgung im östlichen Mittelmeerraum (4. bis 15. Jahrhundert)

13 *Op. cit.* footnote 10.

14 Thyveetil, M. A., Manos, S., Suter J. L. and Coveney, P. V. Use of UKLight as a Fast Network for Grid Infrastructures In Clarke P., Davenhall C., Greenwood C. and Strong M. (Eds.)Lighting the Blue Touchpaper for UK e-Science - Closing Conference of the ESLEA Project. PoS(ESLEA)013 (http://pos.sissa.it//archive/conferences/041/013/ESLEA_013.pdf)

15 Gaffney, V. and Fletcher, R. 2007: Always the Bridesmaid and never the Bride! Arts, Archaeology and the E-Science Agenda. In Clarke P., Davenhall C., Greenwood C. and Strong M. (Eds.) Lighting the Blue Touchpaper for UK e-Science - Closing Conference of the

The concepts underlying scale and the nature of scale effects are therefore likely to remain an issue of considerable academic concern. In disciplines, including the humanities, where behavioural, cognitive or phenomenological issues remain central to our interpretative position, the current awareness of the multivocality of scale will probably increase and become increasingly contentious. Paradoxically, this debate may deepen with increased access to massive computational power. Whilst likely to facilitate more substantive analysis of scale-related data, we may have to be more critical of our own analytical procedures as a consequence. In an imperfect world scale is, indeed, a meagre product of reality.

SCALE AS A CONSIDERATION WITHIN ARCHAEOLOGICAL COMPUTATION

William Kilbride, Glasgow Museums

Introduction

A recent e-Science workshop deliberated various aspects of archaeological computing, using the themes of scale, heterogeneity and standards. Although the themes are significantly linked, scale emerged as perhaps the most intractable of the three themes, in part because the concept carries multiple competing meanings. Although the workshop invited us to consider geospatial computing, the issues discussed were by no means restricted to the cartographic sense of scale.

Scale as Interpretative Practice

Vince Gaffney's introductory paper put the topic of scale and resolution at the centre of the interpretative hermeneutics of the humanities. The constitution of data in the humanities is more complicated than agreement and application on approved metrics to a known sample. For archaeologists and historians, the past can only be understood through a partial and imperfect record which can be assessed at a number of levels: what we chose to use as evidence is almost as important as the theories we think the evidence could support. So scale is an interpretative construct as well as a mathematical transformation. Working at different levels of generalisation implies more than differences in methodology, a point that is all too often lost. Different types of question can be asked, different types of evidence can be assessed, and different assumptions about the world are brought into focus. Routes and communications provide a good example of this overlooked axiom. The Silk Route, which stretched many thousands of miles and operated over many centuries is just not the same as a track through a small valley, even if the latter is a segment of the former. So the tools necessary to understand one may be inadequate or inappropriate to the other. In the digital age, it seems that zooming in or out is trivial: the transformation is simple but the implications are not.

The Scale of Time

Gaffney's astute linkage of time and space was perhaps understated. It is surprising that Einstein's thoughts on relativity have not had a more profound impact on archaeology. If space and time cannot exist independently of each other, then it follows that the whole lexis of spatial science must have a temporal component. Scale is temporal too. We have a well worked series of controls for spatial scale and resolution – the scale bar, the icons, the ratios and so forth – and we are rightly fastidious about their deployment. It is surely alarming that we have no such qualms

about temporal scales nor even the vocabulary in which to describe them. We may use new technology to represent a small area or an entire continent, but unless we know the temporal scales of the map then the representation is intellectually flawed. Using the routes example again, 'time' may represent the time taken to traverse it, or the time taken to establish it, or the routines of life along it. Each of these is different, and moving between them is not trivial. Think for a moment of all those delicately engineered viewsheds that seem less clever when you ask about trees in summer, or about night time. If we are to understand space and time together then we need to develop rapidly a language to describe scales of time.

Such considerations seem petulant – almost curmudgeonly – beside the unrivalled wealth at our disposal. Environmental scanning and remote sensing have advanced so quickly in so many fields that we can now acquire and process geographic data at extraordinary speed and with exquisite precision. The scale of the projects we can now affect are truly breathtaking, and one looks back with curious alarm to the years of data gathering which could now be accomplished in days if not hours. This happy coincidence of massive-scale computing power and ever more exacting measurement is surely to be celebrated. But the point is that technological advances have outstripped and diverged from theoretical ones. More and better archaeology does not necessarily follow from more and better data.

Agent-based modelling

Gaffney's insightful commentary opened the way for two case studies on how issues of scale have contributed to quite different archaeological research projects. Agent-based modelling within known geographies emerged as a particular interest.

The MASS project is led by Tony Wilkinson at Durham University, exploring how different economic, political and environmental phenomena impact the development of small communities in northern Syria. The phenomena in question have different scales of impact: taxation for example may impact on a household level while economic or environmental considerations could have regional implications. Set within an exhaustively studied landscape and supported by numerous contextual sources, the study simulated a number of different communities, and proposed certain types of response to different stimula. For example, the relationship between grain yield and rainfall can be predicted within certain constraints, so the impact additional stress factors such as labour shortages or taxation could be explored. Patterns of economic and demographic development could be modelled in the medium to long term, and the impact on local and regional environments could to some extent be predicted.

Mark Lake from University College London took a step back to describe the challenges, benefits and implications of agent-based modelling in archaeology, providing specific thoughts on how, over a duration, very localised practices can have very large scale impact. A worked example from the mesolithic of Islay supported the presentation, examining how ecological variables impact on settlement processes. In essence a small group of discrete but mutually impacting decisions over a long period produce what looks like a homogenous pattern –a homogeneity that has heterogeneous origins.

Discussion

The shift to a consideration of agent-based modelling was at first surprising. In a workshop about scale, one would have anticipated questions about examining large phenomena through multiple small scale analyses. On first inspection the session

should have been about resource scheduling on distributed computing and excited predictions about what we may be able to achieve in the future – the leitmotif of so many E-science workshops. The caution about theoretical naivety could have been expected, but it was surprising to revisit a decade or more of debate in the aims and outcomes of archaeological computing.

Gaffney's introductory remarks prepared the debate. For example, the iterative processes of agent-based modelling are underpinned by notions of space *and time*. The MASS project evaluated economic patterns over a 100-year range: implying that the scale of the analysis is over several generations rather than a single lifetime or season. Thus, what appears from a spatial perspective as an intimate, localised study is in fact geared towards macro-scale processes.

It is fair to say that opinions are divided about the benefits of agent-based modelling, and that they have been divided for some time. As early as 1970, Doran was predicting that cybernetic analyses of human systems would revolutionise archaeological research[16] – but from almost the same time critics have questioned the assumptions which underpin such research. The 1980s saw speculation on the properties and opportunities of artificial intelligence and expert systems to assist research – machines and programmes that were always just too complicated to become mainstream[17] [18]. By the 1990's ideas of process and system had been replaced with ideas of agency, and a battery of techniques like viewshed and cost-surface had sought to incorporate landscape, experience and meaning into the debate. By that point the debate had resolved to a discussion of environmental determinism and the role of computers as an aid to archaeological interpretation not as an end in itself[19]. The fundamentals of that debate perhaps need restated because some of the conditions that pertained then exist still:

- Computing is not a neutral methodology, and although large sums of public money tend to be invested in computing infrastructure (then it was SERC's Science-Based Archaeology programme, this time it's e-science) the existence of funding is not an intellectual framework;
- We may have ready access to environmental information and tools to process them – pollen samples, terrain models, geological mapping, interpolation algorithms and the like – but over-reliance on them is a danger: there is more to humanity than our relationship to the environment;
- Archaeological evidence is partial and fragmentary and this creates issues of circularity in analyses. It's not always clear whether we are seeking to account for the formation of the archaeological record or to interpret it;
- There can be no tabula rasa. Even an empty landscape can be meaningfully constituted, and the concept of 'year 0' is absurd. Consequently, the edge effects that are recognised from computing are rendered all the more

---

[16] Doran, J. 1970 Systems theory, computer simulations and archaeology. *World Archaeology* 289-298

[17] Huggett, J. 1985: Expert systems in archaeology, in J. Richards and M. Cooper (eds.) *Current Issues in Archaeological Computing* (British Archaeological Reports, Oxford), 123-142.

[18] Gardin, J.C. 1989: Artificial Intelligence and Expert Systems: Case Studies in the Knowledge Domain of Archaeology, Prentice Hall (Sd).

[19] Gaffney V. and M. van Leusen, 1995. Postscript: GIS and environmental determinism: a parallel text. In G. Lock and Z. Stancic (Eds) *Archaeology and Geographic Information Systems*. London, 367-382.

complicated in cultural landscapes. There is an infinite residual from the previous state, and the scope of a study is necessarily too narrow.

What does this mean for Scale as a phenomenon in archaeological computing? The foregoing discussion in fact resolves to a question of the scale of agency, a topic that has been very thoroughly discussed in sociology and anthropology. Is agent-based modelling really just a new implementation of systems theory? Reliance on individual decision making sounds like a move away from generalised, homogenising processes in favour of more discrete and convincing phenomena, but the terms of reference available for such agent-based decisions are very constrained and, by implication, individual agency is compromised. In discussion, the claim was made that such modelling requires implicit assumptions to be made explicit, which is an obvious benefit because it enables others to adopt or criticise the work: such transparency is not expected of other approaches and is never volunteered. Thus it was claimed that agent-based modelling offers a more virtuous paradigm for research. But is this not exactly the positivist doctrine of a New Archaeology which post-processualism killed off a decade ago?

Further thoughts

It is not for the rapporteur to resolve the debate nor even to take particular sides, but a number of observations do seem pertinent.

The discussion revealed genuinely mixed opinions on the benefits or otherwise of agent-based modelling, skepticism on the ability of its proponents fully to declare their implicit assumptions, and doubt as to anyone picking up such an analysis would ever be able to handle the very large audit trail such documentation suggests would be produced. There have always been competing opinions on what constitutes a compelling argument in archaeology. Consequently the difficulty of forming a consensus on agent-based modelling is not a sign of weakness.

It does seem however that the discursive impact of computing, and our understanding of the role of the computer as a tool has moved on. Computing is not neutral and the results of computer-based analyses are no longer taken on their own merits. At its best, modelling is exploratory and experimental: supporting and supported by a range of other approaches. As ever, external forces are often the drivers behind clumsy adaptations of archaeology. The computer gaming industry was discussed in this light, but the same themes can be detected in discussions about the computer graphics in the 1980s and GIS in the 1990s.

It is also worth noting the themes that didn't emerge in the workshop but which perhaps should have. Community engagement with and awareness of the e-Science programme remains stubbornly low. This is perhaps associated with the pressures of the Research Assessment Exercise and uncertainty as to whether credit could be claimed for the development and publication of data and tools. It was also surprising that there was little discussion of the continuing rise of high performance computing. It surely follows that some of our data gathering and processing projects are currently complex but will become trivial as enhanced network capacity and computing power becomes available. Surely we can set out our priorities for the judicious use of scarce research funding with an eye to the future as well as the present.

SESSION 2: HETEROGENEITY

MASHUP OR MESSUP? THE CHOICE IS (NOT) UP TO US

Thomas Elliott, University of North Carolina

I have occasionally described the Barrington Atlas and its companion Map-by-Map Directory as 'a giant spatial index into the scholarly literature for Greek and Roman places.' It was the editorial policy of the Classical Atlas Project (1988-2000), which produced the atlas, to require its scholarly compilers to furnish one or more relevant citations of secondary scholarly literature before a feature could be mapped in the atlas and listed in the directory. Where published secondary work was lacking, primary source citations and (rarely) the authority of named scholars working at the site or in the region were admitted. This practice has obvious benefits for both the editors, and the users, of the atlas. It also hints at the vast and varied universe of information that was searched, sifted, collected, interpreted, synthesized, argued and adjusted in creating the 99 maps and 1,500 pages of supporting data in tabular format. In fact, these tables and maps can be viewed as a regularized, well-structured user interface to an inherently heterogeneous dataset, itself compiled from multiple, differently structured sources that had been created and published for the widest imaginable range of purposes.

Scholars (humanists and scientists alike) work at the interface between heterogeneity and homogenization. We regularly wrestle with complex, chaotic and often contradictory or ambiguous sources, empirical data and prior conclusions in an effort to produce new interpretations that advance knowledge and inform future investigation. When we get it right, new understandings emerge. When we get it wrong, we get *pablum*.

The traditional scholarly article, monograph or reference work in print almost always constitutes a remix of other data and scholarship, or presents findings based on analysis of such a remix. Even empirical datasets created through laboratory or field work arise through interpretative and classificatory processes that produce new order from observed chaos. Standard bibliographic citation, recognized conventions for textual apparatus, explicit invocation of theoretical interpretative positions and exhaustive descriptions of methodology are all effective techniques for signalling - to human readers - underlying heterogeneity and potential discontinuity. Confidence intervals and other statistical measures provide additional qualification and context for some numeric data.

Recent advances in computing, digital culture and scholarly practice are opening up new possibilities and potential pitfalls. In particular, the rising popularity of virtual globe software, neogeographical computing practices and mashups are lowering barriers to 3D visualization as a tool for teaching, research and recreation. The corresponding upsurge in interest is driving the development of easier and more powerful mechanisms for harvesting and aggregating spatially referenced data. Many of these mechanisms are quick and dirty; they bypass the elaborate schemas, protocols and metadata content standards developed by the geospatial computing industry and science funding bodies. Instead, they favour simplicity, economy of expression and lowest-common-denominator web patterns.

Yet - despite a proliferation of data models and encoding formats for feature services, gazetteers, earth browsers and geographical tags - it has proved impossible for us to encode all aspects of our project's legacy dataset (the Barrington Atlas itself) in any single standard schema. At present, we are using (internally) a

'frankenformat' in which the simplest and most useful pieces of various schemas are ganged together to provide the needed data transport. The most obvious shortcoming of this approach is its idiosyncrasy. No one else has existing code that can parse this format, so publishing our data in it would have limited value. For data interchange we have so far relied on various more standard serializations (KML, Atom + GeoRSS), but do so at the cost of 'dumbing down' our data.

It seems to me that one of the more urgent tasks facing archaeologists, historians and other humanists interested in establishing production-level spatial services and methods (or publishing work compatible with them) is the resolving of this data encoding and interchange problem. I am not arguing for the creation of yet another schema or protocol. Rather, I think we must renew efforts to engage with the existing format-and-tools communities to advocate for our needs. We should especially push for the adoption of solutions that can be used, unchanged, across multiple spatial data formats and that, preferably, have already wide use or an active development community. Among the present gaps are:

- Robust methods for communicating uncertainty, accuracy, precision and similar factors (both qualitative and quantitative) in computationally actionable ways;
- Well known and widely implemented conventions for the (carto)graphical interpretation of such indicators in data;
- Similar mechanisms for transmitting and surfacing novel representations of scholarly process or data provenance now becoming possible for born-digital works;
- Flexible and precise ways to associate events (including durations) and subjects (tags, categories) with places and names and communicate these associations;
- Non-idiosyncratic citation formats for primary and secondary sources (in both print and digital form) that communicate roles (attestation, provenance, argumentation or additional information) and that can be easily mashed up with third-party bibliographic and document-delivery services.

To engage successfully with other communities in pursuit of these goals, we must not only attend the same conferences and invite outsiders to meetings convened for the purpose of collaboration, but we must also conduct case studies with real datasets and existing formats, and then publish the results of those experiments. Such work will inevitably involve review of more than data formats. Conceptual models, digital creation processes, editorial workflow and preservation strategies will all undergo evaluation and change. And we must seek more than a series of schemas and associated technical documentation. A body of published best practices, backed up by accessible, exemplary resources and services, is essential.

RAPPORTEUR'S REPORT – HETEROGENEITY

David Wheatley, University of Southampton

Introduction

Rapporteur[20] (derived from French) is used in international and European legal and political contexts to refer to a person appointed by a deliberative body to investigate an issue or a situation, and report back to that body.

Never having been entirely clear what a *rapporteur* is, and having been given no clear guidance by the organisers, I turned to Wikipedia for a definition (above). This may be a useful definition, and it may even be correct although, like all information, I have to make that judgement while considering its origins. In this case, it is helpful to me and so I chose to make use of it to guide me in writing this report. As such, I've opted to assume that my role is to report back to the organizers my impressions of the papers and discussions on heterogeneity. The report therefore consists of my impressions of the three contributions, followed by some personal thoughts that stem from the debate.

Heterogeneity

It is tempting as an archaeological practitioner to regard the heterogeneity of our data as a problem. It seems to make our jobs – whether as researchers, communicators, administrators or fieldworkers – harder to do, and therefore represents something that needs to be controlled, suppressed and discouraged. However, this variety of sources and data types with their different, often highly subtle, semantics is in reality part of the character of our discipline: archaeology requires us to consider elements as diverse as classical texts, pottery reports, radiocarbon determinations, photographs, plans and antiquarian accounts.

The three speakers in the session each illustrated this in different ways, and drew attention to the different responses that ICT professionals have to heterogeneity.

Tom Elliott drew attention to the issues that faced the producers of the Barrington Atlas of the Greek and Roman World as it tries to exploit the undoubted advantages of electronic delivery, with its opportunities for collaborative update and recurring publication. The 'complex chaotic and contradictory' sources of the atlas bring into sharp focus the currently competing philosophies of 'top down' control and 'bottom up' facilitation. The first of these summarizes traditional attempts to control data (through metadata, spatial data infrastructures, data clearinghouses) whose aim is pre-planned interoperability. 'Bottom up' approaches, by contrast, derive from so-called 'Web 2.0' ideas and the growth of 'folksonomies' which allow participants to 'just get on with it' by adding whatever tags and codes they deem relevant. Whatever our personal investment in the first of these, Elliott drew our attention to the fact that the web 2.0 'genie is out of the bottle': the success and growth of neogeographies (Google Earth and its derivative activities similar) means that we return to a world in which we expect to maintain tight control of vocabulary, and where interoperability will be enforced through standards. The outcome for the Barrington Atlas seems to fall somewhere between these extremes, and has been driven by practical necessity rather than ideological or political investment and was described by Elliott as a 'Frankenformat': an internal, task specific mixture of different formats that seem to do the job.

[20] According to Wikepedia (what else!)

By contrast, **Stuart Jeffrey** presented the perspective of those charged with ensuring that heterogeneous archaeological spatial datasets remain accessible to future generations of archaeologists. We must all be familiar now with the difficulty of the task that the Archaeology Data Service (ADS) has set itself, because its success requires us to preserve not just our data but the semantic context of it. Jeffrey drew our attention to the ways that the ADS is approaching spatial datasets such as the ArchSearch facilities, where records within the ADS catalogue (the index to a hugely heterogeneous range of databases from a hugely heterogeneous range of sources) can now be retrieved in a variety of ways, including spatial query and 'clicky map' interfaces. The ADS appear to be wrestling with the omni-present tension between standards and diversity (see discussion in the following sections), and the response has been (unsurprisingly, given that the ADS has to deal with complex issues relating to data currency, accuracy, precision, ownership and copyright) a focus on 'top down' technologies for data control, interoperability and cross-searching (such as Z39.50) rather than to move swiftly into the 'bottom up' technologies that appear to be bypassing much of this effort. More surprisingly, perhaps, Jeffrey provided insight into the political tensions that exist over the relative level control and influence that major players (notably English Heritage and the ADS) should be allowed to have over the future of spatial data provision. It also became apparent that the ADS policy of not archiving and providing software is clearly a limitation on some possible solutions.

From the other side, in a sense, of the archaeological research process **Leif Isaksen** presented some thoughts about Ptolemy's *Geography*, how Ptolemy dealt with heterogeneity and uncertainty in his own sources and then how the 'invisible parameters' of spatial data might be computationally defined. It seems that Ptolemy understood that the level of accuracy of his spatial data was less than he seems to superficially claim. A bit like the famous axiom about advertising[21], this leads to the knowledge that *some* of Ptolemy's information is quite precise (even if only relatively so) but it is not possible to know *which* parts. Isaksen compares this historical precision issue with modern malpractices in spatial referencing, such as the frequent misuse of the imprecise forms of OS National Grid referencing within GIS and with the rise of neogeographies, he argues, these spatial referencing issues may become more widespread and significant. Isaksen proposes a new conceptual framework for spatial representation which could provide useful 'axes' for categorizing map data so that it may be handled computationally, allowing provision of spatial data in a more intelligent manner while also compelling the compiler of spatial data to express its epistemic limits.

Heterogeneity?

Debate at the meeting was at times predictable, and at times surprising. Unsurprisingly, there was attention given to the perceived problems of 'folksonomies', with discussion of how (or indeed if) to control information. As an example, the georeferencing of photographs using panoramia was cited as an example of uncontrolled structuring of data, permitting 'tourist nonsense' to invade otherwise structured datasets. Less immediately obvious (and popular) was the suggestion that, were we all to write our archaeologies according to Jean Claude Gardin's inference chain methodology then sufficient structure would exist to allow us to derive ontologies and establish complex interrelated datasets.

21 Department store pioneer John Wanamaker is credited with saying "I know I waste half the money I spend on advertising. The problem is, I don't know which half."

Underlying these discussions, in my view, is the tension that must exist in any attempt to impose order on the heterogenous chaos that is archaeological data. This is that 'knowledge is power'[22] and that control of digital information is a particularly potent expression of that power. In the digital world, therefore, it may be inevitable that connecting our information to other researchers and other communities requires us to give up some of this power and allow others to act in ways that we don't control, perhaps don't approve of and certainly can't predict. Our desire to corral our data into increasingly complex structures so that their metadata can be effectively mated with other metadata is apparently well motivated, but may have at its heart a reluctance to surrender enough of our knowledge to others. That collective desire to impose order has been expressed most clearly within our discipline through the foundation and growth of the ADS. We need to recognise that while solutions to the delivery of complex spatial datasets to the research community can be facilitated by the existence of agencies like the ADS, they may at the same time be hampered by the political and institutional issues that arise from the existence of those agencies.

Our need to impose order onto heterogeneity carries opportunities and risks. Clearly, there are advantages to the formulation of a conceptual framework for spatial datasets in that this may allow re-use of spatial data in ways that avoid the kinds of precision issues that Ptolemy so usefully implements. How these could be made transparent, or operationalized in the context of, say, Google Earth is less apparent though and if we seek to insist that all those providing spatial data comply with a pre-determined framework we may end up prohibiting the rapid development of neogeographies which, for all their chaotic and at times nonsensical connections, have successfully converted digital provision into a melting pot of ideas and opportunities.

In my own view these tensions will continue for the foreseeable future, with most of us ultimately accepting some middle ground. Meanwhile, we can help ourselves to deal with heterogeneity in a number of practical ways. Firstly, we can throw more of our data away: it is, of course, arrogant to believe that we have a right to dig and interpret archaeology without ensuring that our data is available to future generations but it is, in my view, equally arrogant to assume that future generations should be obliged to preserve every last cogitation of each of our research careers. We need to make an objective assessment of what future generations of archaeologists may reasonably expect us to keep, and discard the rest. Secondly, we need to worry less about the existence of 'less rigorous' sources of information and 'wrong' information. Information is consumed in an intelligent way, and users are able to distinguish, say, Wikipedia entries from peer-reviewed journals. We also need to worry less that 'users' will do inappropriate things with our data. They will, however hard we try to protect ourselves with metadata and ontologies. While we (rightly) want to inform future potential users of our datasets about their origins and limitations, ultimately what is done with it in the future is not our responsibility, and in any case many not be as much of a problem as we believe (after all, inappropriate things may only be inappropriate from our own perspective). There are, in reality, no 'naïve users' and there is no 'unambiguous information' and users (both 'us' the archaeologists and 'them' the consumers of archaeological information) are increasingly smart enough to know that today's world is composed of heterogeneous information, increasing amounts of it spatially referenced.

For me, it follows from these vague thoughts that recent concern over dealing with the heterogeneity in our spatial datasets through ever complex spatial data

22 This equally axiomatic quote is hardly recent, originating from the author and philosopher Sir Francis Bacon (1597, *Religious Meditations, Of Heresies*).

infrastructures, ontologies and interoperability mechanisms can actually get in the way of the practical task of delivering our (often imperfect, incorrect, imprecise) information, of just getting on with it in the way that neogeographers have decided to do. If we remain focused, as the Barrington Atlas has, on the research and communication tasks and resist the temptation to become ever more sophisticated in our design of systems for describing, cataloguing (and controlling) data at the expense of a focus on research and communication then we may find that heterogeneity is a joyous and beneficial thing rather than a problem. We need to remember at all times that it is this variety and diversity – the heterogeneity – of archaeological data that makes the past so interesting to investigate.

SESSION 3: STANDARDS AND METADATA

STANDARDS AND METADATA IN GEOSPATIAL DATA: SOLUTIONS TO OTHER PEOPLE'S PROBLEMS?

William Kilbride, Glasgow Museums

The topic of standards and metadata (and standards of metadata) in geospatial data are widely discussed by better commentators than this one, and it remains an area of active development. In this sense, the workshop will present us with latest thinking and novel solutions to well known problems. Rather than summarizing - and doing considerable violence - to the work of the OGC and others with respect to standards, this short introductory paper focuses on three specific aspects of standards development. Purposefully contentious, it will be proposed that standards (and metadata) are characteristically extraneous: they are adopted not for their own sake nor for any internal logic. Instead, data standards can be categorized as solutions to other people's problems.

In the first instance it will be noted that the role of standards development and promotion and training comes with a discursive price tag: that being 'in the know' or 'on the inside' provides an ineluctable professional authority which is not easily assailed or assessed. Secondly we are told that data standards promote data sharing. Experience suggests that the presumption in favour of data sharing within the academic community and the presumed role of the academic community in the context of a national spatial data infrastructure is more contentious than policy documents might lead us to suppose. Institutional infrastructure and legal impediments are predisposed to disrupt that aspiration to data sharing, threatening to render the standards hypothetical. Finally we are told to adopt standards for the longue duree: that preservation requires documentation. This is undeniably true but the complexities of geospatial data and the relative immaturity of the operational standards for trusted repositories means that conventional, archive-based models will be sorely tested to provide the sort of long term support that we quite evidently need.

This short essay is intended to be provocative. It is most certainly not the opinion of Glasgow Museums, nor of the JISC Geospatial Working Group. It is not even necessarily of the opinion of the author.

Reduction to language

Knowledge, we are told, is power. This could be stated more elegantly and precisely: literacy is discursive. It establishes, maintains and disguises relationships of dependency and autonomy. In that sense there is no naïve literacy and the infrastructure associated with the maintenance and monitoring of approved forms of literacy are not trivial. This is most apparent in historical contexts where access to

literacy and thus knowledge was carefully managed, and more importantly where relationships of dependency were created and made self-evident, such as the relationship between literate clergy and illiterate laity.

Working on the assumption that modern information technology is simply the latest innovation in the long history of literacy and language, we should expect to find these same discursive realities hidden in our own information technology: who is deciding what can and cannot be known? Who is placed in a relationship of dependency from which they cannot easily escape? This issue can be explored in part through standards development. The need and desire to share information means, for example, an impetus towards controlled vocabulary in the humanities, and expectations about the coherence of geographical description. But the naming of things matters in the humanities, so the adoption of a shared language – someone else's language – risks violence to the subject of study. Phenomenologically speaking, it's not possible to put your foot into the same river once. Nor is this concern with language and meaning confined to the philosophical ramblings of humanists: cartography is exhaustingly political. So will the search for shared protocols and semantic interoperability wreck the interpretative project of the humanities? This is a moot point – it's not clear that the external reality of the world can be reduced to language at all, let alone someone else's language. Stepping away from the solipsistic cliff edge, and assuming that the world can be contained in language, it should be clear that the there is more to the promotion and adoption of vocabulary controls and spatial syntax than might first appear. They are a solution to someone else's problem and the problem is getting you to do what they want. We need to have a great big row about standards.

Intellectual Rights and Wrongs

From the recondite planes of linguistics to the quotidian bustle of our offices and institutions, it's not even clear that data sharing is a universal virtue. There is still more lip service than web service. Policy documents proudly proclaim the merits of open access and so we come to expect that data sets will be available to us after an appropriate interval. A few brave souls are good enough to provide instant access. But as the commercial value of the data increases so the ease of access declines. This is especially true of geospatial data. It is customary to criticize mapping agencies for their reluctance to sacrifice their one major asset on the altar of open access, but the reality is that many of our institutions play this game too. It is easy to trace the Ordnance Survey's caution to a Thatcherite agenda of fiscal independence, but the research community is under the same pressures. On one hand universities are expected to collaborate in an open and sharing environment, on the other they are expected to compete and develop IPR-based business plans which turn ideas into 'third leg funding'. Data sharing is fine so long as the integrity of data and the profits of the host institution are not compromised.

Problems of intellectual property rights are not insurmountable. The will to succeed and trust in colleagues are powerful forces that time and again mean the issues resolve themselves. But the inadequacy of current legislation is in stark contrast to the relative sophistication of what could be achieved. This is especially true when we consider the complexity of derived data sets in which it is no longer clear who owns what, and therefore difficult to be confident in what can and cannot be re-supplied. Clearer documentation and demarcation of sources is an obvious solution, but the quantities of metadata required and the complexities of licensing suggest that this is likely to be a temporary solution. The development of a national spatial data infrastructure implies data sharing: and for this to happen there has to be a simple and well-understood protocol for who is responsible for what. It's unlikely that tools

like Creative Commons could become the norm in the spatial domain – but the sort of empowering clarity which it implies is at least worth the protracted effort.

Institutional aspirations for repositories present another, and somewhat unexpected, challenge to the free flow of geospatial data. Still in their infancy, the stock-in-trade of repositories are digital research papers: theses, journal articles and the like. The scope of many repositories maps conveniently onto submissions for the Research Assessment Exercise, and somewhat inconveniently onto the range of digital outputs that research produce and consume. Few if any are able to support the technical.

Towards the digital heritage service

Intellectual property rights and institutional policies leads us seamlessly into strategies for the preservation of geospatial data. Institutional repositories are characteristically not designed for long-term preservation and the complexity of intellectual property laws inhibit much reasonable short term action for long term gain.

The conventional wisdom of digital preservation envisages a trusted digital repository managing sets of files with appropriate administrative, technical and representational metadata to enable and ensure independent utility. Setting aside for the moment the technical complexity involved in rendering geospatial data – one should assume the designated community and technical sophistication of the archive managers can deal with such issues – the implications of so much derived data should become obvious. It is unreasonable to expect that a single repository will be able to manage all the components. Granted the Open Archival Information System allows for different functions within a single system to be distributed (an OAIS can span various institutions as AHDS has shown). But it does not envisage that the same functions be replicated in multiple agencies. Even allowing for this discrepancy the advent of change-only-updates and live sensor feeds confuses the argument even more. It is hard to imagine a single repository being able to take responsibility for the long term preservation of an integrated GIS project: it seems much more likely that a number of agencies will have to work together. For this to happen there needs to be a matrix of mutually understood and compatible responsibilities with each agent continually assessing the performance and viability of the partners. Mutually managed and distributed curation sounds attractive but the tools and standards for this sort of preservation are still only in draft. This sort of ‘preservation in situ’ is metaphorically closer to heritage management than archives, so is superficially attractive to archaeology and the historic environment, but the similarities are more imagined than real. Perhaps a family of law and practice could be imagined with scheduled ancient data sets of national importance protected as such. But if we can’t fix immediate and basic problems like permanent identifiers the prospects are not good in the medium term.

Conclusion

Standards and standards developers really need to address three issues. At a basic level we need to know why so few people interested in standards development, and decide if we are happy with that. Do we need a much bigger effort to involve the whole community or is it appropriate to leave it in the hands of special interests and technocrats (like me). The standards for information exchange are well developed and are arguably way ahead of institutional policy. This means that they risk becoming hypothetical. Arguably we should stop developing standards and start redeveloping our institutions. Finally the long term for geospatial data is far from secure, and with a proliferation of services and agencies, is likely to become less

secure in time (not more). The national spatial data infrastructure wants a future then it needs to get a history.

STANDARDS AND METADATA – RAPPORTEUR'S REPORT

Stephen Stead, Paveprime Ltd.

The session had two speakers followed by a discussion and concluded with a rapporteur presentation. This short paper tries to document this last presentation, and reflects only the rapporteurs view of the rest of the session.

William Kilbride presented a lavishly illustrated and thought-provoking analysis of the role of standards in our community. He explained that standards are an external and artificial imposition on our practice and data, and as such are intrinsically political. It also means that the setting of standards allows the manifestation of political power: they that control the standards control the world! This perhaps goes part way to explaining the plethora of data standards in the heritage sector. He continued by pointing out that standards are slow moving whereas technology is swift and nimble. This is to be expected of course as the two have different driving imperatives. Standards are intended to stabilise and add continuity whereas technology is driven by the need to innovate and open new markets.

Kilbride noted that Standards are not neutral; they perpetuate a particular morale, theoretical and social view. As long as this is recognised all is well, but a mistaken faith in the benign, disinterested neutrality of a particular standard is tantamount to the disengagement of all critical faculties. One feature of all standards frequently confused by their developers is the difference between the identity of something and its name(s). The bar beside the river in Ljubljana is the same whether it is called The Black Cat; Kavarna Maček; Parmova 41 or 'the bar beside the river in Ljubljana'. The separation of these things is all to frequently not done at all, or is not made explicit and so is difficult to understand when data is shared. In order to facilitate sharing, one needs three things: content, consistency and trust. It does not matter how big, or what quality, the data is as long as it is internally consistent and we can trust its associated documentation. Finally, he noted that digital preservation needs us, as a discipline, to undertake risk assessments for our data, and this will inevitably require new models of risk and data use.

Guy McGarva gave a comprehensive guide to the EDINA national data centre spatial data and services[23]. He stressed the role that standards play in EDINA's operations. In particular they provide for:

- Interoperability
- Universal access
- Device independence
- Architectural integrity
- Preservation

He also stressed the role of standards in providing delivery formats supporting all of these aspects. In the future, EDINA intends to continue to provide discovery and access points through a variety of metadata standards, but acknowledges that this is problematic due to the large number of metadata elements that spatial data can have.

[23] http://www.edina.ac.uk (last accessed 14/2/2008)

The discussion concentrated on access and availability, but made a number of other thought provoking points. The debate about how to provide access to data churned over the questions of who should be allowed what sort of access. The basic issues concern the resources required for maintenance and how those are funded, who gets to use the data and what rights accrue to the data originator in the products of re-users efforts. A useful distinction was attempted between Database Right and Copyright but with little concrete conclusion. This is an area upon which a wider and informed debate is required. It was noted that there a number of barriers to sharing data. These include the legal restrictions of copyright, originators' reluctance to share data that they are still using in their as yet unpublished research, prejudice and the fact that the data may be below par (i.e. no one wants to reveal that their data is crap!). It was generally felt that much of this could be overcome with clear, explicit statements of where data originated; a sort of citation for data.

A further point was that the advent of the World Wide Web allows a wide range of alternative producers of data (see also the preceding section). Some content creators produce well documented and researched complementary data sets, while others produce a plethora of poorly researched, terribly documented or 'alternative' data. Some of this clearly has an agenda and can be grouped under the 'nutters' heading. Of course such data need not be incorporated into our research, but it complicates the task of due diligence (i.e. checking that all relevant sources have been consulted), and makes dissemination more difficult as there are so many 'tree like' objects obscuring the woods.

The second point was that we must not lose sight of the fact that ad hoc or local standards, that support particular research questions, are to be welcomed. Attempts to stifle such innovation produce stagnation within the discipline. However, if an ad hoc solution is not explicitly documented then any work using it is automatically lost upon completion of the project, even if great care is taken to provide long term preservation of the data. Data without documentation is a waste of shelf space. Finally alternative spaces were discussed. Most spatial data is tied to a model of the earth as a globe (or ablate spheroid). However the possibility exists for us to use other non-Cartesian spaces to situate our research: for example to use mental maps, non-linear space or song-lines to frame our research area. Currently we have few examples of this and best practice has yet to evolve.

Conclusion

The world of standards can usefully be broken into three levels. At the bottom, and least influenced by the Heritage sector, are **technical standards**. These include file formats, communication and computer standards and universal data value declarations (e.g. ISO country codes). In the middle are **content standards**. This is the area within which the Heritage sector can exercise innovation, and must do so to stop stagnation. Finally there are the **metadata** standards that support resource discovery and integration. It is at this last level that CIDOC Conceptual Reference Model (CRM) sits[24]. It provides us with the mechanism to allow interoperability or integration between the multitudes of research orientated content standards. Every project should routinely publish a mapping of its data structure to the CRM along with the rest of the documentation of method and practice used in the project.

The pragmatic result of any work on Standards should be that our data is consistent, our process documented and our documentation explicit. If we achieve that then our

[24] http://cidoc.ics.forth.gr (last accessed 14/2/2008)

work will survive as the profession as a whole will be able to reuse its results. If we fail in any part of this then our data is damned and can be safely deleted at the end of the project as it is no use to man nor beast.

APPENDIX I: Participants

Dr Elton Barker, University of Oxford
Dr Gabriel Bodard, King's College London
Dr Stephen Boyd Davis, Middlesex University
Dr Stuart Dunn, King's College London
Dr Thomas Elliott, University of North Carolina
Prof Vince Gaffney, University of Birmingham
Mr Christopher Green, University of Leicester
Mr Leif Isaksen, Oxford Archaeology (now University of Southampton)
Dr Stuart Jeffrey, University of York
Dr William Kilbride, Glasgow Museums
Dr Mark Lake, University College London
Dr Annalisa Marzano, University of Oxford
Mr Guy McGarva, Edina
Mr John Onyango, Glasgow School of Art
Dr Torseten Reimer, King's College London
Mr Stephen Stead, Paveprime Ltd
Mr Bruce Stenning
Dr Jenny Ure, University of Edinburgh
Dr Hafed Walda, King's College London
Dr David Wheatley, University of Southampton
Prof Tony Wilkinson, University of Durham

APPENDIX 2: Presentations

- 'Finding Your Way to the Map: the Challenges of Delivering Geospatial Data for Archaeologists' - Stuart Jeffrey, University of York
- 'An Infrastructure View on Space and Place' - James Reid, EDINA
- 'Aliens and Spatio-Temporal Things: a GIScience Perspective' - Femke Reitsma, University of Edinburgh
- 'Scale: A Meagre Product of Reality' - Vince Gaffney, University of Birmingham
- 'The MASS Project' - Tony Wilkinson, University of Durham
- 'Agent-Based Modelling: A Question of Scale' - Mark Lake, UCL.
- Mash-Up or Mess-Up: It's (Not) Up To Us' - Thomas Elliott, University of North Carolina
- The Challenges of Delivering Geospatial Data for Archaeologists' - Stuart Jeffrey, University of York
- Ptolemy's Error: Truths and Falsehoods in Heterogeneous Spatial Data' - Leif Isaksen, Oxford Archaeology

- ‘Standards and Metadata in Geospatial Data: Solutions for Other People’s Problems?’ William Kilbride, Glasgow Museums

- ‘Experiences at EDINA: The Role of Standards and Metadata in Spatial Data Infrastructures’ - Guy McGarva, EDINA